Characterizing the NP-PSPACE Gap in the Satisfiability Problem for Modal Logic*

Joseph Y. Halpern
Computer Science Department
Cornell University, U.S.A.
e-mail: halpern@cs.cornell.edu

Leandro Chaves Rêgo †
Statistics Department
Federal University of Pernambuco, Brazil
e-mail: leandro@de.ufpe.br

Abstract

There has been a great deal of work on characterizing the complexity of the satisfiability and validity problem for modal logics. In particular, Ladner showed that the satisfiability problem for all logics between $\mathbf{K}$ and $\mathbf{S4}$ is *PSPACE*-hard, while for $\mathbf{S5}$ it is *NP*-complete. We show that it is *negative introspection*, the axiom $\neg Kp \Rightarrow K\neg Kp$, that causes the gap: if we add this axiom to any modal logic between $\mathbf{K}$ and $\mathbf{S4}$, then the satisfiability problem becomes *NP*-complete. Indeed, the satisfiability problem is *NP*-complete for any modal logic that includes the negative introspection axiom.

Keywords: Modal Logic, Complexity, Satisfiability Problem, Negative Introspection, Euclidean Property.

---

*This work was supported in part by NSF under grants CTC-0208535, ITR-0325453, and IIS-0534064, by ONR under grants N00014-00-1-03-41 and N00014-01-10-511, and by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the ONR under grant N00014-01-1-0795. The second author was also supported in part by a scholarship from the Brazilian Government through the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). A preliminary version of this paper appeared in the 20th International Joint Conference on Artificial Intelligence (IJCAI 2007).

†Most of this work was done while the author was at the School of Electrical and Computer Engineering at Cornell University, USA. Mail should be sent to this author at Rua Muniz Tavares 25, apt. 902, Jaqueira, Recife-PE, Brazil, Zip: 52050-170. Phone: +55-81-2126-8421.

1 Introduction

There has been a great deal of work on characterizing the complexity of the satisfiability and validity problem for modal logics (see [7; 9; 14; 15] for some examples). In particular, Ladner [9] showed that the validity (and satisfiability) problem for every modal logic between $\mathbf{K}$ and $\mathbf{S4}$ is *PSPACE*-hard; and is *PSPACE*-complete for the modal logics $\mathbf{K}$, $\mathbf{T}$, and $\mathbf{S4}$. He also showed that the satisfiability problem for $\mathbf{S5}$ is *NP*-complete.

What causes the gap between *NP* and *PSPACE* here? We show that, in a precise sense, it is the negative introspection axiom: $\neg K\varphi \Rightarrow K\neg K\varphi$. It easily follows from Ladner's proof of *PSPACE*-hardness that for any modal logic L between $\mathbf{K}$ and $\mathbf{S4}$, there exists a family of formulas φ_n, all consistent with L such that such that $|\varphi_n| = O(n)$ but the smallest Kripke structure satisfying φ has at least 2^n states (where $|\varphi|$ is the length of φ viewed as a string of symbols). By way of contrast, we show that for all of the (infinitely many) modal logics L containing $\mathbf{K5}$ (that is, every modal logic containing the axiom K—$K\varphi \wedge K(\varphi \Rightarrow \psi) \Rightarrow K\psi$—and the negative introspection axiom, which has traditionally been called axiom 5), if a formula φ is consistent with L, then it is satisfiable in a Kripke structure of size linear in $|\varphi|$. Using this result and a characterization of the set of finite structures consistent with a logic L containing $\mathbf{K5}$ due to Nagle and Thomason [12], we can show that the consistency (i.e., satisfiability) problem for L is *NP*-complete. Thus, roughly speaking, adding negative introspection to any logic between $\mathbf{K}$ and $\mathbf{S4}$ lowers the complexity from *PSPACE*-hard to *NP*-complete.

The fact that the consistency problem for specific modal logics containing $\mathbf{K5}$ is *NP*-complete has been observed before. As we said, Ladner already proved it for $\mathbf{S5}$; an easy modification (see [6]) gives the result for $\mathbf{KD45}$ and $\mathbf{K45}$.[1] That the negative introspection axiom plays a significant role has also been observed before; indeed, Nagle [11] shows that every formula φ consistent with a normal modal logic[2] L containing $\mathbf{K5}$ has a finite model (indeed, a model exponential in $|\varphi|$) and using that, shows that the provability problem for every logic L between $\mathbf{K}$ and $\mathbf{S5}$ is decidable; Nagle and Thomason [12] extend Nagle's result to all logics containing $\mathbf{K5}$ not just normal logics. Despite all this prior work and the fact that our result follows from a relatively straightforward combination of results of Nagle and Thomason and Ladner's techniques for proving that the consistency problem for $\mathbf{S5}$ is *NP*-complete, our result seems to be new, and is somewhat surprising (at least to us!).

The rest of the paper is organized as follows. In the next section, we review standard notions from modal logic and the key results of Nagle and Thomason [12] that we use. In Section 3, we prove the main result of the paper. We discuss related work in Section 4.

2 Modal Logic: A Brief Review

We briefly review basic modal logic, introducing the notation used in the statement and proof of our result. The syntax of the modal logic is as follows: formulas are formed by starting with

[1] Nguyen [13] also claims the result for $\mathbf{K5}$, referencing Ladner. While the result is certainly true for $\mathbf{K5}$, it is not immediate from Ladner's argument.

[2] A modal logic is *normal* if it satisfies the generalization rule RN: from φ infer $K\varphi$.

a set $\Phi = \{p, q, \ldots\}$ of primitive propositions, and then closing off under conjunction ($\wedge$), negation ($\neg$), and the modal operator K. Call the resulting language $\mathcal{L}_1^K(\Phi)$. (We often omit the Φ if it is clear from context or does not play a significant role.) As usual, we define $\varphi \vee \psi$ and $\varphi \Rightarrow \psi$ as abbreviations of $\neg(\neg\varphi \wedge \neg\psi)$ and $\neg\varphi \vee \psi$, respectively. The intended interpretation of $K\varphi$ varies depending on the context. It typically has been interpreted as knowledge, as belief, or as necessity. Under the epistemic interpretation, $K\varphi$ is read as "the agent *knows* φ"; under the necessity interpretation, $K\varphi$ can be read "φ is necessarily true".

The standard approach to giving semantics to formulas in $\mathcal{L}_1^K(\Phi)$ is by means of Kripke structures. A tuple $F = (S, \mathcal{K})$ is a *(Kripke) frame* if S is a set of states and $\mathcal{K}$ is a binary relation on S. A *situation* is a pair (F, s), where $F = (S, \mathcal{K})$ is a frame and $s \in S$. A tuple $M = (S, \mathcal{K}, \pi)$ is a *Kripke structure (over* Φ*)* if $(S, \mathcal{K})$ is a frame and $\pi : S \times \Phi \rightarrow \{\mathbf{true}, \mathbf{false}\}$ is an *interpretation (on* S*)* that determines which primitive propositions are true at each state. Intuitively, $(s, t) \in \mathcal{K}$ if, in state s, state t is considered possible (by the agent, if we are thinking of K as representing an agent's knowledge or belief). For convenience, we define $\mathcal{K}(s) = \{t : (s, t) \in \mathcal{K}\}$.

Depending on the desired interpretation of the formula $K\varphi$, a number of conditions may be imposed on the binary relation $\mathcal{K}$. $\mathcal{K}$ is *reflexive* if for all $s \in S$, $(s, s) \in \mathcal{K}$; it is *transitive* if for all $s, t, u \in S$, if $(s, t) \in \mathcal{K}$ and $(t, u) \in \mathcal{K}$, then $(s, u) \in \mathcal{K}$; it is *serial* if for all $s \in S$ there exists $t \in S$ such that $(s, t) \in \mathcal{K}$; it is *Euclidean* iff for all $s, t, u \in S$, if $(s, t) \in \mathcal{K}$ and $(s, u) \in \mathcal{K}$ then $(t, u) \in \mathcal{K}$. We use the superscripts r, e, t and s to indicate that the $\mathcal{K}$ relation is restricted to being reflexive, Euclidean, transitive, and serial, respectively. Thus, for example, $\mathcal{S}^{rt}$ is the class of all situations where the $\mathcal{K}$ relation is reflexive and transitive.

We write $(M, s) \models \varphi$ if φ is true at state s in the Kripke structure M. The truth relation is defined inductively as follows:

$$
\begin{aligned}
&(M, s) \models p, \text{ for } p \in \Phi, \text{ if } \pi(s, p) = \mathbf{true} \\
&(M, s) \models \neg\varphi \text{ if } (M, s) \not\models \varphi \\
&(M, s) \models \varphi \wedge \psi \text{ if } (M, s) \models \varphi \text{ and } (M, s) \models \psi \\
&(M, s) \models K\varphi \text{ if } (M, t) \models \varphi \text{ for all } t \in \mathcal{K}(s)
\end{aligned}
$$

A formula φ is said to be *satisfiable in Kripke structure* M if there exists $s \in S$ such that $(M, s) \models \varphi$; φ is said to be *valid in* M, written $M \models \varphi$, if $(M, s) \models \varphi$ for all $s \in S$. A formula is *satisfiable* (resp., *valid*) *in a class* $\mathcal{N}$ of Kripke structures if it is satisfiable in some Kripke structure in $\mathcal{N}$ (resp., valid in all Kripke structures in $\mathcal{N}$). There are analogous definitions for situations. A Kripke structure $M = (S, \mathcal{K}, \pi)$ is *based on* a frame $F = (S', \mathcal{K}')$ if $S' = S$ and $\mathcal{K}' = \mathcal{K}$. The formula φ is *valid in* situation (F, s), written $(F, s) \models \varphi$, where $F = (S, \mathcal{K})$ and $s \in S$, if $(M, s) \models \varphi$ for all Kripke structure M based on F.

Modal logics are typically characterized by axiom systems. Consider the following axioms and inference rules, all of which have been well-studied in the literature [3; 4; 6]. (We use the traditional names for the axioms and rules of inference here.) These are actually *axiom schemes* and *inference schemes*; we consider all instances of these schemes.

Prop. All tautologies of propositional calculus

K. $(K\varphi \wedge K(\varphi \Rightarrow \psi)) \Rightarrow K\psi$ (Distribution Axiom)

T. $K\varphi \Rightarrow \varphi$ (Knowledge Axiom)

4. $K\varphi \Rightarrow KK\varphi$ (Positive Introspection Axiom)

5. $\neg K\varphi \Rightarrow K\neg K\varphi$ (Negative Introspection Axiom)

D. $\neg K(\mathit{false})$ (Consistency Axiom)

MP. From φ and $\varphi \Rightarrow \psi$ infer ψ (Modus Ponens)

RN. From φ infer $K\varphi$ (Knowledge Generalization)

The standard modal logics are characterized by some subset of the axioms above. All are taken to include Prop, MP, and RN; they are then named by the other axioms. For example, $\mathbf{K5}$ consists of all the formulas that are provable using Prop, K, 5, MP, and RN; we can similarly define other systems such as $\mathbf{KD45}$ or $\mathbf{KT5}$. $\mathbf{KT}$ has traditionally been called $\mathbf{T}$; $\mathbf{KT4}$ has traditionally been called $\mathbf{S4}$; and $\mathbf{KT45}$ has traditionally been called $\mathbf{S5}$.

For the purposes of this paper, we take a *modal logic* L to be any collection of formulas that contains all instances of Prop and is closed under modus ponens (MP) and substitution, so that if φ is a formula in L and p is a primitive proposition, then $\varphi[p/\psi] \in L$, where $\varphi[p/\psi]$ is the result of replacing all instances of p in φ by ψ. A logic is *normal* if it contains all instances of the axiom K and is closed under the inference rule RN. In terms of this notation, Ladner [9] showed that if L is a normal modal logic between $\mathbf{K}$ and $\mathbf{S4}$ (since we are identifying a modal logic with a set of formulas here, that just means that $\mathbf{K} \subseteq L \subseteq \mathbf{S4}$), then determining if $\varphi \in L$ is *PSPACE*-hard. (Of course, if we think of a modal logic as being characterized by an axiom system, then $\varphi \in L$ iff φ is provable from the axioms characterizing L.) We say that φ is *consistent with* L if $\neg\varphi \notin L$. Since consistency is just the dual of provability, it follows from Ladner's result that testing consistency is *PSPACE*-hard for every normal logic between $\mathbf{K}$ and $\mathbf{S4}$. Ladner's proof actually shows more: the proof holds without change for non-normal logics, and it shows that some formulas consistent with logics between $\mathbf{K}$ and $\mathbf{S4}$ are satisfiable only in large models. More precisely, it shows the following:

Theorem 2.1: [9]

(a) *Checking consistency is PSPACE complete for every logic between* $\mathbf{K}$ *and* $\mathbf{S4}$ *(even non-normal logics).*

(b) *For every logic* L *between* $\mathbf{K}$ *and* $\mathbf{S4}$*, there exists a family of formulas* φ_n^L*,* $n = 1, 2, 3, \ldots$*, such that (i) for all* n*,* φ_n^L *is consistent with* L*, (ii) there exists a constant* d *such that* $|\varphi_n^L| \leq dn$*, (iii) the smallest Kripke structure that satisfies* φ *has at least* 2^n *states.*

There is a well-known correspondence between properties of the $\mathcal{K}$ relation and axioms: reflexivity corresponds to T, transitivity corresponds to 4, the Euclidean property corresponds to 5, and the serial property corresponds to D. This correspondence is made precise in the following well-known theorem (see, for example, [6]).

Theorem 2.2: *Let $\mathcal{C}$ be a (possibly empty) subset of $\{T, 4, 5, D\}$ and let C be the corresponding subset of $\{r, t, e, s\}$. Then $\{Prop, K, MP, RN\} \cup \mathcal{C}$ is a sound and complete axiomatization of the language $\mathcal{L}_1^K(\Phi)$ with respect to $\mathcal{S}^C(\Phi)$.*[3]

Given a modal logic L, let $\mathcal{S}^L$ consist of all situations (F, s) such that $\varphi \in L$ implies that $(F, s) \models \varphi$. An immediate consequence of Theorem 2.2 is that $\mathcal{S}^e$, the situations where the $\mathcal{K}$ relation is Euclidean, is a subset of $\mathcal{S}^{\mathbf{K5}}$.

Nagle and Thomason [12] provide a useful semantic characterization of all logics that contain $\mathbf{K5}$. We review the relevant details here. Consider all the finite situations $((S, \mathcal{K}), s)$ such that either

1. S is the disjoint union of S_1, S_2, and $\{s\}$ and $\mathcal{K} = (\{s\} \times S_1) \cup ((S_1 \cup S_2) \times (S_1 \cup S_2))$, where $S_2 = \emptyset$ if $S_1 = \emptyset$; or
2. $\mathcal{K} = S \times S$.

Using (a slight variant of) Nagle and Thomason's notation, let $\mathcal{S}_{m,n}$, with $m \geq 1$ and $n \geq 0$ or $(m, n) = (0, 0)$, denote all situations of the first type where $|S_1| = m$ and $|S_2| = n$, and let $\mathcal{S}_{m,-1}$ denote all situations of the second type where $|S| = m$. (The reason for taking -1 to be the second subscript for situations of the second type will become clearer below.) It is immediate that all situations in $\mathcal{S}_{m,n}$ for fixed m and n are isomorphic—they differ only in the names assigned to states. Thus, the same formulas are valid in any two situations in $\mathcal{S}_{m,n}$. Moreover, it is easy to check that the $\mathcal{K}$ relation in each of the situations above in Euclidean, so each of these situations is in $\mathcal{S}^{\mathbf{K5}}$. It is well known that the situations in $\mathcal{S}_{m,-1}$ are all in $\mathcal{S}^{\mathbf{K5}}$ and the situations in $\mathcal{S}_{m,-1} \cup \mathcal{S}_{m,0}$ are all in $\mathcal{S}^{\mathbf{KD45}}$. In fact, $\mathbf{S5}$ (resp., $\mathbf{KD45}$) is sound and complete with respect to the situations in $\mathcal{S}_{m,-1}$ (resp., $\mathcal{S}_{m,-1} \cup \mathcal{S}_{m,0}$). Nagle and Thomason show that much more is true. Let $\mathcal{T}^L = (\cup\{\mathcal{S}_{m,n} : m \geq 1, n \geq -1 \text{ or } (m, n) = (0, 0)\}) \cap \mathcal{S}^L$.

Theorem 2.3: [12] *For every logic L containing $\mathbf{K5}$, L is sound and complete with respect to the situations in $\mathcal{T}^L$.*

The key result of this paper shows that if a formula φ is consistent with a logic L containing $\mathbf{K5}$, then there exists m, n, a Kripke structure $M = (S, \mathcal{K}, \pi)$, and a state $s \in S$ such that

[3] We remark that soundness and completeness is usually stated with respect to the appropriate class $\mathcal{M}^C$ of structures, rather than the class $\mathcal{S}^C$ of situations. However, the same proof applies without change to show completeness with respect to $\mathcal{S}^C$, and using $\mathcal{S}^C$ allows us to relate this result to our later results. While for normal logics it suffices to consider only validity with respect to structures, for non-normal logics, we need to consider validity with respect to situations.

$((S,\mathcal{K}),s) \in \mathcal{S}_{m,n}$, $\mathcal{S}_{m,n} \subseteq \mathcal{T}^L$, and $m+n < |\varphi|$. That is, if φ is satisfiable at all, it is satisfiable in a situation with a number of states that is linear in $|\varphi|$.

One more observation made by Nagle and Thomason will be important in the sequel.

Definition 2.4: A *p-morphism* (short for *pseudo-epimorphism*) from situation $((S',\mathcal{K}'),s')$ to situation $((S,\mathcal{K}),s)$ is a function $f : S' \rightarrow S$ such that

- $f(s') = s$;
- if $(s_1,s_2) \in \mathcal{K}'$, then $(f(s_1),f(s_2)) \in \mathcal{K}$;
- if $(f(s_1),s_3) \in \mathcal{K}$, then there exists some $s_2 \in S'$ such that $(s_1,s_2) \in \mathcal{K}'$ and $f(s_2) = s_3$.

∎

This notion of p-morphism of situations is a variant of standard notions of p-morphism of frames and structures [3]. It is well known that if there is a p-morphism from one structure to another, then the two structures satisfy the same formulas. An analogous result holds for situations.

Theorem 2.5: *If there is a p-morphism from situation (F',s') to (F,s), then for every modal logic L, if $(F',s') \in \mathcal{S}^L$ then $(F,s) \in \mathcal{S}^L$.*

Proof: Suppose that $F = (S,\mathcal{K})$, $F' = (S',\mathcal{K}')$, f is a p-morphism from (F',s') to (F,s), and $(F',s') \in \mathcal{S}^L$. We want to show that $(F,s) \in \mathcal{S}^L$. Let Φ be the set of primitive propositions. Given an interpretation π on S, define an interpretation $\pi' : S' \times \Phi \rightarrow \{\mathbf{true},\mathbf{false}\}$ on S' by taking $\pi'(t,p) = \pi(f(t),p)$ for all $t \in S'$ and $p \in \Phi$. We now show by induction on the structure of formulas that for all states $t \in S'$ and all formulas φ, we have $(F',\pi',t) \models \varphi$ iff $(F,\pi,f(t)) \models \varphi$. This is a standard argument [3]; we repeat it here for completeness.

The base case follows immediately from the definition of π'. For conjunctions and negations the argument is immediate from the induction hypothesis. Finally, if φ is of the form $K\varphi'$, first suppose that $(F',\pi',t) \models K\varphi'$. We want to show that $(F,\pi,f(t)) \models K\varphi'$. So suppose that $(f(t),u) \in \mathcal{K}$. Since f is a p-morphism, then there exists $u' \in S'$ such that $(t,u') \in \mathcal{K}'$ and $f(u') = u$. Since $(F',\pi',t) \models K\varphi'$, it must be the case that $(F',\pi',u') \models \varphi'$. By the induction hypothesis, it follows that $(F,\pi,u) \models \varphi'$. Since this argument applies to all u such that $(f(t),u) \in \mathcal{K}$, it follows that $(F,\pi,f(t)) \models K_i\varphi'$. For the opposite implication, suppose that $(F,\pi,f(t)) \models K\varphi'$. We want to show that $(F',\pi',t) \models K\varphi'$. If $(t,u) \in \mathcal{K}'$ then, since f is a p-morphism, $(f(t),f(u)) \in \mathcal{K}$. Since $(F,\pi,f(t)) \models K\varphi'$, it follows that $(F,\pi,f(u)) \models \varphi'$. By the induction hypothesis, $(F',\pi',u) \models \varphi'$. It follows that $(F',\pi',t) \models K\varphi'$.

To complete the argument, suppose by way of contradiction that $\varphi \in L$ and $(F,s) \not\models \varphi$. Then there exists an interpretation π such that $(F,\pi,s) \models \neg\varphi$. Since $f(s') = s$, by the argument above, there exists an interpretation π' on S' such that $(F',\pi',s') \models \neg\varphi$, contradicting the assumption that $(F',s') \in \mathcal{S}^L$. ∎

Now consider a partial order on pairs of numbers, so that $(m, n) \leq (m', n')$ iff $m \leq m'$ and $n \leq n'$. Nagle and Thomason observed that if $(F, s) \in \mathcal{S}_{m,n}$, $(F', s') \in \mathcal{S}_{m',n'}$, and $(1, -1) \leq (m, n) \leq (m', n')$, then there is an obvious p-morphism from (F', s') to (F, s): if $F = (S, \mathcal{K})$, $S = S_1 \cup S_2$, $F' = (S', \mathcal{K}')$, $S' = S'_1 \cup S'_2$ (where S_i and S'_i for $i = 1, 2$ are as in the definition of $\mathcal{S}_{m,n}$), then define $f : S' \to S$ so that $f(s') = s$, f maps S'_1 onto S_1, and, if $S_2 \neq \emptyset$, then f maps S'_2 onto S_2; otherwise, f maps S'_2 to S_1 arbitrarily. The following result (which motivates the subscript -1 in $\mathcal{S}_{m,-1}$) is immediate from this observation and Theorem 2.5.

Theorem 2.6: *If $(F, s) \in \mathcal{S}_{m,n}$, $(F', s') \in \mathcal{S}_{m',n'}$, and $(1, -1) \leq (m, n) \leq (m', n')$, then for every modal logic L, if $(F', s') \in \mathcal{T}^L$ then $(F, s) \in \mathcal{T}^L$.*

3 The Main Results

We can now state our key technical result.

Theorem 3.1: *If L is a modal logic containing* **K5** *and $\neg\varphi \notin L$, then there exist m, n such that $m + n < |\varphi|$, a situation $(F, s) \in \mathcal{S}^L \cap \mathcal{S}_{m,n}$, and structure M based on F such that $(M, s) \models \varphi$.*

Proof: By Theorem 2.3, if $\neg\varphi \notin L$, there is a situation $(F', s_0) \in \mathcal{T}^L$ such that $(F', s_0) \not\models \neg\varphi$. Thus, there exists a Kripke structure M' based on F' such that $(M', s_0) \models \varphi$. Suppose that $F' \in \mathcal{S}_{m',n'}$. If $m' + n' < |\varphi|$, we are done, so suppose that $m' + n' \geq |\varphi|$. Note that this means $m' \geq 1$. We now construct a a situation $(F, s) \in \mathcal{S}_{m.n}$ such that $(1, -1) \leq (m, n) \leq (m', n')$, $m + n < |\varphi|$, and $(M, s) \models \varphi$ for some Kripke structure based on F. This gives the desired result. The construction of M is similar in spirit to Ladner's [9] proof of the analogous result for the case of **S5**.

Let C_1 be the set of subformulas of φ of the form $K\psi$ such that $(M', s_0) \models \neg K\psi$, and let C_2 be the set of subformulas of φ of the form $K\psi$ such that $KK\psi$ is a subformula of φ and $(M', s_0) \models \neg KK\psi \wedge K\psi$. (We remark that it is not hard to show that if $\mathcal{K}$ is either reflexive or transitive, then $C_2 = \emptyset$.)

Suppose that $M' = (S', \mathcal{K}', \pi')$. For each formula $K\psi \in C_1$, there must exist a state $s_\psi^{C_1} \in \mathcal{K}'(s_0)$ such that $(M', s_\psi^{C_1}) \models \neg\psi$. Note that if $C_1 \neq \emptyset$ then $\mathcal{K}'(s_0) \neq \emptyset$. Define $I(s_0) = \{s_0\}$ if $s_0 \in \mathcal{K}'(s_0)$, and $I(s_0) = \emptyset$ otherwise. Let $S_1 = \{s_\psi^{C_1} : K\psi \in C_1\} \cup I(s_0)$. Note that $S_1 \subseteq \mathcal{K}'(s_0) = S'_1$, so $|S_1| \leq |S'_1|$. If $K\psi \in C_2$ then $KK\psi \in C_1$, so there must exist a state $s_\psi^{C_2} \in \mathcal{K}'(s_{K\psi}^{C_1})$ such that $(M', s_\psi^{C_2}) \models \neg\psi$. Moreover, since $(M', s_0) \models K\psi$, it must be the case that $s_\psi^{C_2} \notin \mathcal{K}'(s_0)$. Let $S_2 = \{s_\psi^{C_2} : K\psi \in C_2\}$. By construction, $S_2 \subseteq S'_2$, so $|S_2| \leq |S'_2|$, and S_1 and S_2 are disjoint. Moreover, if $S_1 = \emptyset$, then $C_1 = \emptyset$, so $C_2 = \emptyset$ and $S_2 = \emptyset$.

Let $S = \{s_0\} \cup S_1 \cup S_2$. Define the binary relation $\mathcal{K}$ on S by taking $\mathcal{K}(s_0) = S_1$ and $\mathcal{K}(t) = S_1 \cup S_2$ for $t \in S_1 \cup S_2$. To show that $\mathcal{K}$ is well defined, we must show that (a)

$s_0 \notin S_2$ and (b) if $s_0 \in S_1$, then $S_2 = \emptyset$. For (a), suppose by way of contradiction that $s_0 \in S_2$. Thus, there exists $s \in S_1$ such that $s_0 \in \mathcal{K}'(s)$. By the Euclidean property, it follows that $s_0 \in \mathcal{K}'(s_0)$, a contradiction since S_2 is disjoint from $\mathcal{K}'(s_0)$. For (b), note that if $s_0 \in S_1$, then $s_0 \in \mathcal{K}'(s_0)$. It is easy to see that if $s, s' \in \mathcal{K}'(s_0)$, then $\mathcal{K}'(s) = \mathcal{K}'(s')$. For if $s, s' \in \mathcal{K}'(s_0)$ then, by the Euclidean property, $s' \in \mathcal{K}'(s)$. Thus, if $t \in \mathcal{K}'(s)$, another application of the Euclidean property shows that $t \in \mathcal{K}'(s')$. Hence, $\mathcal{K}'(s') \subseteq \mathcal{K}'(s)$. A symmetric argument gives equality. But now suppose that $t \in S_2$. Then, as we have observed, there exists some $s \in S_1$ such that $t \in \mathcal{K}'(s) - \mathcal{K}'(s_0)$. But if $s_0 \in S_1$, then $\mathcal{K}'(s) - \mathcal{K}'(s_0) = \emptyset$. Thus, $S_2 = \emptyset$ if $s_0 \in S_1$.

A similar argument shows that $\mathcal{K}$ is the restriction of $\mathcal{K}'$ to S. For clearly S_2 is disjoint from $\mathcal{K}'(s_0)$, so $\mathcal{K}(s_0) = \mathcal{K}'(s_0) \cap S$. Now suppose that $s \in S_1 \cup S_2$. It is easy to see that there exists some $s' \in S_1$ such that $s \in \mathcal{K}'(s')$. This is clear by construction if $s \in S_2$. And if $s \in S_1$, then $s \in \mathcal{K}'(s_0)$ and, by the Euclidean property, $s \in \mathcal{K}'(s)$. If $t \in S_1 \cup S_2$, we want to show that $t \in \mathcal{K}'(s)$. Again, there exists some t' such that $t' \in S_1$ and $t \in \mathcal{K}'(t')$. Since $s', t' \in \mathcal{K}'(s_0)$, by the Euclidean property, $s' \in \mathcal{K}'(t')$. Since $s', t \in \mathcal{K}'(t')$, the Euclidean property implies that $t \in \mathcal{K}'(s')$. Since $s, t \in \mathcal{K}'(s')$, yet another application of the Euclidean property shows that $t \in \mathcal{K}'(s)$. Thus, $\mathcal{K}(s) \subseteq \mathcal{K}'(s) \cap S$. To prove equality suppose that $t \in \mathcal{K}'(s) \cap S$. If $t \in S_1 \cup S_2$, then by definition $t \in \mathcal{K}(s)$. If $t = s_0$, then by the Euclidean property it follows that $s_0 \in \mathcal{K}'(s_0)$, so $s_0 \in S_1 \subseteq \mathcal{K}(s)$. Thus, $t \in \mathcal{K}(s)$, as desired.

Let $M = (S, \mathcal{K}, \pi)$, where π is the restriction of π' to $\{s_0\} \cup S_1 \cup S_2$. It is well known [6] (and easy to prove by induction on φ) that there are at most $|\varphi|$ subformulas of φ. Since C_1 and C_2 are disjoint sets of subformulas of φ, all of the form $K\psi$, and at least one subformula of φ is a primitive proposition (and thus not of the form $K\psi$), it must be the case that $|C_1| + |C_2| \leq |\varphi| - 1$, giving us the desired bound on the number of states.

We now show that for all states $s \in S$ and for all subformulas ψ of φ (including φ itself), $(M, s) \models \psi$ iff $(M', s) \models \psi$. The proof proceeds by induction on the structure of φ. The only nontrivial case is when ψ is of the form $K\psi'$. If $(M', s) \models K\psi'$, then $(M', t) \models \psi'$ for all $t \in \mathcal{K}'(t)$. Since $\mathcal{K}$ is the restriction of $\mathcal{K}'$ to S, this implies that $(M', t) \models \psi'$ for all $t \in \mathcal{K}(s)$. Thus, by the induction hypothesis, $(M, t) \models \psi'$ for all $t \in \mathcal{K}(s)$; that is, $(M, s) \models K\psi'$. For the converse, suppose that $(M', s) \models \neg K\psi'$. If it is also the case that $(M', s_0) \models \neg K\psi'$, then $K\psi' \in C_1$. By the construction of M and the induction hypothesis, $(M, s_{\psi'}^{C_1}) \models \neg\psi'$. Thus, $(M, s) \models \neg K\psi'$. If $(M', s_0) \models K\psi'$, then standard arguments using the fact that $\mathcal{K}'$ is Euclidean can be used to show $(M', s_0) \models \neg KK\psi'$. Thus, $K\psi' \in C_2$, and $(M, s_{\psi'}^{C_2}) \models \neg\psi'$ by the induction hypothesis. Again, it follows that $(M, s) \models \neg K\psi'$.

By construction, $(F, s) \in \mathcal{S}_{m,n}$, where $m = |S_1|$ and $n = |S_2|$. We have already observed that $m + n < |\varphi|$, $|S_1| \leq |S_1'|$, and $|S_2| \leq |S_2'|$. Thus, $(m, n) \leq (m', n')$. It follows from Theorem 2.6 that $(F, s) \in \mathcal{T}^L \subseteq \mathcal{S}^L$. This completes the proof. ∎

The idea for showing that the consistency problem for a logic L that contains $\mathbf{K5}$ is *NP*-complete is straightforward. Given a formula φ that we want to show is consistent with L, we simply guess a frame $F = (S, \mathcal{K})$, structure M based on F, and state $s \in S$ such that $(F, s) \in \mathcal{S}_{m,n}$ with $m + n < |\varphi|$, and verify that $(M, s) \models \varphi$ and $\mathcal{S}_{m,n} \subseteq \mathcal{T}^L$. Verifying that

$(M,s) \models \varphi$ is the *model-checking problem*. It is well known that this can be done in time polynomial in the number of states of M, which in this case is linear in $|\varphi|$. So it remains to show that, given a logic L containing $\mathbf{K5}$, checking whether $\mathcal{S}_{m,n} \subseteq \mathcal{T}^L$ can be done efficiently. This follows from observations made by Nagle and Thomason [12] showing that that, although $\mathcal{T}^L$ may include $\mathcal{S}_{m',n'}$ for infinitely many pairs (m',n'), $\mathcal{T}^L$ has a finite representation that makes it easy to check whether $\mathcal{S}_{m.n} \subseteq \mathcal{T}^L$.[4]

Say that (m,n) is a *maximal index* of $\mathcal{T}^L$ if $m \geq 1$, $\mathcal{S}_{m,n} \subseteq \mathcal{T}^L$, and it is not the case that $\mathcal{S}_{m',n'} \subseteq \mathcal{T}^L$ for some (m',n') with $(m,n) < (m',n')$. It is easy to see that $\mathcal{T}^L$ can have at most finitely many maximal indices. Indeed, if (m,n) is a maximal index, then there can be at most $m+n-1$ maximal indices, for if (m',n') is another maximal index, then either $m' < m$ or $n' < n$ (for otherwise $(m,n) \leq (m',n')$, contradicting the maximality of (m,n)). Say that $m \geq 1$ is an *infinitary first index* of $\mathcal{T}^L$ if $\mathcal{S}_{m,n} \subseteq \mathcal{T}^L$ for all $n \geq -1$. Similarly, say that $n \geq -1$ is an *infinitary second index* of $\mathcal{T}^L$ if $\mathcal{S}_{m,n} \subseteq \mathcal{T}^L$ for all $m \geq 1$. Note that it follows from Theorem 2.6 that if $(1,-1) \leq (m,n) \leq (m',n')$, then if m' is an infinitary first index of $\mathcal{T}^L$, then so is m, and if n' is an infinitary second index of $\mathcal{T}^L$, then so is n. Suppose that m^* is the largest infinitary first index of $\mathcal{T}^L$ and n^* is the largest infinitary second index of $\mathcal{T}^L$, where we take $m^* = n^* = \infty$ if all first indices are infinitary (or, equivalently, if all second indices are infinitary), we take $m^* = -1$ if no first indices are infinitary, and we take $n^* = -2$ if no second indices are infinitary. It follows from all this that $\mathcal{T}^L$ can be represented by the tuple $(i, m^*, n^*, (m_1,n_1), \ldots, (m_k,n_k))$, where

- i is 1 if $\mathcal{S}_{0,0} \in \mathcal{T}^L$, and 0 otherwise;
- m^* is the largest infinitary first index;
- n^* is the largest infinitary second index; and
- $((m_1,n_1), \ldots, (m_k,n_k)$ are the maximal indices.

Given this representation of $\mathcal{T}^L$, it is immediate that $\mathcal{S}_{m,n} \subseteq \mathcal{T}^L$ iff one of the following conditions holds:

- $(m,n) = (0,0)$ and $i = 1$;
- $1 \leq m \leq m^*$;
- $-1 \leq n \leq n^*$; or
- $(m,n) \leq (m_k,n_k)$.

[4] The representation that we are about to give is similar in spirit to, although not the same as, that of Nagle and Thomason. (We find ours both easier to present and easier to work with.)

We can assume that the algorithm for checking whether a formula is consistent with L is "hard-wired" with this description of L. It follows that only a constant number of checks (independent of φ) are required to see if $\mathcal{S}_{m,n} \subseteq \mathcal{T}^L$.[5]

Putting all this together, we get our main result.

Theorem 3.2: *For all logics L containing* $\mathbf{K5}$*, checking whether φ is consistent with L is an NP-complete problem.*

We can actually improve Theorem 3.2 slightly. In Theorem 3.2, the logic L is viewed as fixed; the algorithm gets as input just the formula φ. We now show that, given as input a logic L containing K5 and a formula φ, it is *NP*-complete to decide if φ is sastifiable in L. We need to be a little careful here; the logic L consists of an infinite number of formulas, so we must present it in an appropriate way. One way to do this is simply to describe L as above, by a tuple of the form $(i, m^*, n^*, (m_1, n_1), \ldots, (m_k, n_k))$. With this representation, the result clearly holds, since it is easy to check, after guessing a situation $S_{m,n}$ that satisfies φ, whether it is in L. We use a slightly different representation, but one which quickly leads to the same result. As shown by Nagle and Thomason [12], each logic L containing K5 is finitely axiomatizable; thus, we describe L by giving as input its axiomatization. In fact, the axiomatization, which we now describe, closely follows the finite representation of L given above.

For $m \geq 0$, let σ_m be the formula

$$\bigwedge_{i=1}^{m+1} \neg K \neg p_i \Rightarrow \bigvee_{i=1}^{m+1} \bigvee_{j=i+1}^{m+1} \neg K \neg (p_i \wedge p_j),$$

(where $p_1, \ldots, p_{m+1}$ are distinct primitive propositions). Note that if $m = 0$, then the right-hand side of the implication in σ_0 is the empty disjunction, which we identify with the formula *false*. It easily follows that σ_0 is equivalent to $K \neg p_1$. Intuitively, σ_m is valid in situation (F, s) if there are at most m states considered possible at s. Since there are at most m states, the formulas $p_1, \ldots, p_{m+1}$ cannot all be true at different states; there must be some state where two of these formulas are true. (It is easy to see that σ_0, i.e., $K \neg p_1$, is valid in (F, s) iff K*false* is valid in (F, s).)

Similarly, for $m \geq 0$, let τ_m be the formula

$$\bigwedge_{i=1}^{m+1} \neg K K \neg p_i \wedge \bigwedge_{i=1}^{m+1} K \neg p_i \Rightarrow \bigvee_{i=1}^{m+1} \bigvee_{j=i+1}^{m+1} \neg K K \neg (p_i \wedge p_j).$$

It is straightforward to check that τ_0 is equivalent to $K \neg p_1 \Rightarrow K K \neg p_1$. Finally, we define τ_{-1} to be the formula $Kp \Rightarrow p$, $\sigma_\infty = \tau_\infty =$ *true*, and $\sigma_{-1} = \tau_{-2} =$ *false*.

The following lemma is straightforward to check.

[5]Here we have implicitly assumed that checking whether inequalities such as $(m, n) \leq (m', n')$ hold can be done in one time step. If we assume instead that it requires time logarithmic in the inputs, then checking if $\mathcal{S}_{m,n} \subseteq \mathcal{T}^L$ requires time logarithmic in $m + n$, since we can take all of $m^*, n^*, m_1, \ldots m_k, n_1, \ldots, n_k$ to be constants.

Lemma 3.3: *Suppose that* $(F, s) \in \mathcal{S}_{m,n}$ *for some* m, n *with* $m \geq 1$*,* $n \geq -1$ *or* $(m, n) = (0, 0)$*:*

(a) If $k \geq 0$*, then* $(F, s) \models \sigma_k$ *iff* $0 \leq m \leq k$*.*

(b) If $k \geq -1$*, then* $(F, s) \models \tau_k$ *iff* $-1 \leq n \leq k$*.*

It easily follows from Lemma 3.3 that if L is characterized by the tuple

$$R = (0, m^*, n^*, (m_1, n_1), \ldots, (m_k, n_k)),$$

then L is characterized by the axiom

$$\varphi_R = \sigma_{m^*} \vee \tau_{n^*} \vee (\sigma_{m_1} \wedge \tau_{n_1}) \vee \ldots \vee (\sigma_{m_k} \wedge \tau_{n_k})$$

(in addition to the axioms K and 5, and the rules of inference MP and RN).

If L is characterized by the tuple $R = (1, m^*, n^*, (m_1, n_1), \ldots, (m_k, n_k))$, then φ_R has the additional disjunct σ_0.[6]

Theorem 3.4: *Given as input a logic* L *containing* $\mathbf{K5}$ *(where, if* L *is characterized by the tuple* R*, then the input is actually the formula* φ_R*) and a formula* φ*, the problem of deciding whether* φ *is consistent with* L *is NP-complete.*

Proof: The argument is essentially identical to that of Theorem 3.2. We simply guess a frame (F, s) in $\mathcal{S}_{m,n}$ for some m, n with $m + n < |\varphi|$ and an interpretation π and check that $(F, \pi, s) \models \varphi$ and that $(F, s) \models \varphi_R$. The key point is that checking whether $(F, s) \models \varphi_R$ does not require checking that $(F, \pi', s) \models \varphi_R$ for all interpretations π', since the validity of φ_R depends only on m and n. ∎

4 Discussion and Related Work

We have shown that, in a precise sense, adding the negative introspection axiom pushes the complexity of a logic between $\mathbf{K}$ and $\mathbf{S4}$ down from *PSPACE*-hard to *NP*-complete. This is not the only attempt to pin down the *NP-PSPACE* gap and to understand the effect of the negative introspection axiom. We discuss some of the related work here.

A number of results showing that large classes of logics have an *NP*-complete satisfiability problem have been proved recently. For example, Litak and Wolter [10] show that the satisfiability for all finitely axiomatizatble tense logics of linear time is *NP*-complete, and Bezhanishvili and Hodkinson [2] show that every normal modal logic that properly extends $\mathbf{S5^2}$ (where $\mathbf{S5^2}$ is the modal logic that contains two modal operators K_1 and K_2, each of which satisfies the

[6] Because our representation of L is somewhat different than that of Nagle and Thomason, our axiom is somewhat different, although similar in spirit.

axioms and rules of inference of $\mathbf{S5}$ as well as the axiom $K_1K_2p \Leftrightarrow K_2K_1p$) has a satisfiability problem that is *NP*-complete. Perhaps the most closely related result is that of Hemaspaandra [14], who showed that the consistency problem for any normal logic containing $\mathbf{S4}.3$ is also *NP*-complete. $\mathbf{S4}.3$ is the logic that results from adding the following axiom, known in the literature as D1, to $\mathbf{S4}$:

D1. $K(K\varphi \Rightarrow \psi) \vee K(K\psi \Rightarrow \varphi)$

D1 is characterized by the *connectedness* property: it is valid in a situation $((S, \mathcal{K}), s)$ if for all $s_1, s_2, s_3 \in S$, if $(s_1, s_2) \in \mathcal{K}$ and $(s_1, s_3) \in \mathcal{K}$, then either $(s_2, s_3) \in \mathcal{K}$ or $(s_3, s_2) \in \mathcal{K}$. Note that connectedness is somewhat weaker than the Euclidean property; the latter would require that *both* (s_2, s_3) and (s_3, s_2) be in $\mathcal{K}$.

As it stands, our result is incomparable to Hemspaandra's. To make the relationship clearer, we can restate her result as saying that the consistency property for any normal logic that contains $\mathbf{K}$ and the axioms T, 4, and D1 is *NP*-complete. We do not require either 4 or T for our result. However, although the Euclidean property does not imply either transitivity or reflexivity, it does imply *secondary reflexivity* and *secondary transitivity*. That is, if $\mathcal{K}$ satisfies the Euclidean property, then for all states s_1, s_2, s_3, s_4, if $(s_1, s_2) \in \mathcal{K}$, then $(s_2, s_2) \in \mathcal{K}$ and if (s_2, s_3) and $(s_3, s_4) \in \mathcal{K}$, then $(s_2, s_4) \in \mathcal{K}$; roughly speaking, reflexivity and transitivity hold for all states s_2 in the range of $\mathcal{K}$. Secondary reflexivity and secondary transitivity can be captured by the following two axioms:

T′. $K(K\varphi \Rightarrow \varphi)$

4′. $K(K\varphi \Rightarrow KK\varphi)$

Both T′ and 4′ follow from 5, and thus both are sound in any logic that extends $\mathbf{K5}$. Clearly T′ and 4′ also both hold in any logic that extends $\mathbf{S4}.3$, since $\mathbf{S4}.3$ contains T, 4, and the inference rule RN. We conjecture that the consistency property for every logic that extends $\mathbf{K}$ and includes the axioms T′, 4′, and D1 is *NP*-complete. If this result were true, it would generalize both our result and Spaan's result.

Vardi [15] used a difference approach to understand the semantics, rather than relational semantics. This allowed him to consider logics that do not satisfy the K axiom. He showed that some of these logics have a consistency problem that is *NP*-complete (for example, the minimal normal logic, which characterized by Prop, MP, and RN), while others are *PSPACE*-hard. In particular, he showed that adding the axiom $K\varphi \wedge K\psi \Rightarrow K(\varphi \wedge \psi)$ (which is valid in $\mathbf{K}$) to Prop, MP, and RN gives a logic that is *PSPACE*-hard. He then conjectured that this ability to "combine" information is what leads to *PSPACE*-hardness. However, this conjecture has been shown to be false. There are logics that lack this axiom and, nevertheless, the consistency problem for these logics is *PSPACE*-complete (see [1] for a recent discussion and pointers to the relevant literature).

All the results for this paper are for single-agent logics. Halpern and Moses [7] showed that the consistency problem for a logic with two modal operators K_1 and K_2, each of which

satisfies the $\mathbf{S5}$ axioms, is *PSPACE*-complete. Indeed, it is easy to see that if K_i satisfies the axioms of L_i for some normal modal logic L_i containing $\mathbf{K5}$, then the consistency problem for the logic that includes K_1 and K_2 must be *PSPACE*-hard. This actually follows immediately from Ladner's [9] result; then it is easy to see that K_1K_2, viewed as a single operator, satisfies the axioms of $\mathbf{K}$. We conjecture that this result continues to hold even for non-normal logics.

We have shown that somewhat similar results hold when we add awareness to the logic (in the spirit of Fagin and Halpern [5]), but allow awareness of unawareness [8]. In the single-agent case, if the K operator satisfies the axioms K, 5, and some (possibly empty) subset of $\{T, 4\}$, then the validity problem for the logic is decidable; on other hand, if K does not satisfy 5, then the validity problem for the logic is undecidable. With at least two agents, the validity problem is undecidable no matter which subset of axioms K satisfies. We conjecture that, more generally, if the K operator satisfies the axioms of any logic L containing $\mathbf{K5}$, the logic of awareness of unawareness is decidable, while if K satisfies the axioms of any logic between $\mathbf{K}$ and $\mathbf{S4}$, the logic is undecidable.

All these results strongly suggest that there is something about the Euclidean property (or, equivalently, the negative introspection axiom) that simplifies things. However, they do not quite make precise exactly what that something is. More generally, it may be worth understanding more deeply what is about properties of the $\mathcal{K}$ relation that makes things easy or hard. We leave this problem for future work.

Acknowledgments

We thank the IJCAI reviewers for useful comments and for catching an error in a previous version of the paper, and Frank Wolter for suggesting that we add Theorem 3.4.

References

[1] M. ALLEN, *Complexity results for logics of local reasoning and inconsistent belief*, in Theoretical Aspects of Rationality and Knowledge: Proc. Tenth Conference (TARK 2005), 2005, pp. 92–108.

[2] N. BEZHANISHVILI AND I. M. HODKINSON, *All normal extensions of S5-squared are finitely axiomatizable*, Studia Logica, 78 (2004), pp. 443–457.

[3] P. BLACKBURN, M. DE RIJKE, AND Y. VENEMA, *Modal Logic*, Cambridge Tracts in Theoretical Computer Science, No. 53, Cambridge University Press, Cambridge, U.K., 2001.

[4] B. F. CHELLAS, *Modal Logic*, Cambridge University Press, Cambridge, U.K., 1980.

[5] R. FAGIN AND J. Y. HALPERN, *Belief, awareness, and limited reasoning*, Artificial Intelligence, 34 (1988), pp. 39–76.

[6] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi, *Reasoning about Knowledge*, MIT Press, Cambridge, Mass., 1995. A revised paperback edition was published in 2003.

[7] J. Y. Halpern and Y. Moses, *A guide to completeness and complexity for modal logics of knowledge and belief*, Artificial Intelligence, 54 (1992), pp. 319–379.

[8] J. Y. Halpern and L. C. Rêgo, *Reasoning about knowledge of unawareness*, in Principles of Knowledge Representation and Reasoning: Proc. Tenth International Conference (KR '06), 2006, pp. 6–13. Full version available at arxiv.org/cs.LO/0603020.

[9] R. E. Ladner, *The computational complexity of provability in systems of modal propositional logic*, SIAM Journal on Computing, 6 (1977), pp. 467–480.

[10] T. Litak and F. Wolter, *All finitely axiomatizable tense logics of linear time flows are coNP-complete*, Studia Logica, 81 (2005), pp. 153–165.

[11] M. C. Nagle, *The decidability of normal K5 logics*, Journal of Symbolic Logic, 46 (1981), pp. 319–328.

[12] M. C. Nagle and S. K. Thomason, *The extensions of the modal logic K5*, Journal of Symbolic Logic, 50 (1985), pp. 102–109.

[13] L. A. Nguyen, *On the complexity of fragments of modal logics*, in Advances in Modal Logic, Vol. 5, R. Schmidt, I. Pratt-Hartman, M. Reynolds, and H. Wansing, eds., 2005, pp. 249–268.

[14] E. Spaan, *Complexity of modal logics*, PhD thesis, University of Amsterdam, 1993.

[15] M. Y. Vardi, *On the complexity of epistemic reasoning*, in Proc. 4th IEEE Symp. on Logic in Computer Science, 1989, pp. 243–252.